JVI Journals.ASM.org

Unexpected Maintenance of Hepatitis C Viral Diversity following Liver Transplantation

Rebecca R. Gray,[a] Samantha L. Strickland,[b] Nazle M. Veras,[b] Maureen M. Goodenow,[b,c] Oliver G. Pybus,[a] Stanley M. Lemon,[d] Michael W. Fried,[e] David R. Nelson,[f,g] and Marco Salemi[b,h]

Department of Zoology, University of Oxford, Oxford, United Kingdom[a]; Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, Florida, USA[b]; Florida Center for AIDS Research, University of Florida, Gainesville, Florida, USA[c]; Lineberger Comprehensive Cancer Center and the Division of Infectious Diseases, Department of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA[d]; Division of Gastroenterology and Hepatology, School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA[e]; Clinical and Translational Science Institute, University of Florida, Gainesville, Florida, USA[f]; Department of Medicine, University of Florida, Gainesville, Florida, USA[g]; and Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA[h]

Chronic hepatitis C virus (HCV) infection can lead to liver cirrhosis in up to 20% of individuals, often requiring liver transplantation. Although the new liver is known to be rapidly reinfected, the dynamics and source of the reinfecting virus(es) are unclear, resulting in some confusion concerning the relationship between clinical outcome and viral characteristics. To clarify the dynamics of liver reinfection, longitudinal serum viral samples from 10 transplant patients were studied. Part of the E1/E2 region was sequenced, and advanced phylogenetic analysis methods were used in a multiparameter analysis to determine the history and ancestry of reinfecting lineages. Our results demonstrated the complexity of HCV evolutionary dynamics after liver transplantation, in which a large diverse population of viruses is transmitted and maintained for months to years. As many as 30 independent lineages in a single patient were found to reinfect the new liver. Several later posttransplant lineages were more closely related to older pretransplant viruses than to viruses detected immediately after transplantation. Although our data are consistent with a number of interpretations, the persistence of high viral genetic variation over long periods of time requires an active mechanism. We discuss possible scenarios, including frequency-dependent selection or variation in selective pressure among viral subpopulations, i.e., the population structure. The latter hypothesis, if correct, could have relevance to the success of newer direct-acting antiviral therapies.

The hepatitis C virus (HCV; family *Flaviviridae*, genus *Hepacivirus*) infects more than 180 million people worldwide and is a leading global cause of liver disease and cancer (3, 26). Since individuals can remain asymptomatic for decades, the true prevalence is potentially much greater than current estimates of ~3% of the world population (50). At present, no vaccine is available, and pharmacological treatment is only moderately successful, particularly for the most prevalent subtypes circulating in the United States and Europe (3), due in part to a prolonged asymptomatic phase of chronic infection that hinders early identification of HCV transmission among individuals. Although the use of direct-acting antiviral drugs (including two recently licensed protease inhibitors) offer improved sustained antiviral response rates (29), drug treatments will not be successful in all patients (47), and antiviral resistance is likely to play a significant role in treatment failure (21). Thus, chronic HCV infection remains a major public health concern.

Liver cirrhosis develops in up to 20% of HCV-infected individuals, who will eventually require a liver transplant (3). However, the new liver is infected within minutes following transplant (16), serum viral load increases 10- to 20-fold relative to pretransplant levels (12), and the clinical course of disease is accelerated (11). Characteristics of the infecting virus and its anatomical source(s) are unknown, reflecting the lack of a practical and realistic small animal model (30) and challenges in culturing the virus *in vitro*.

The evolutionary rate of HCV within an infected host is on the order of 10^{-3} nucleotide substitutions/site/year (1, 17), resulting in rapid accumulation of mutations over time and circulation of variable viral populations. High variability permits detailed investigation of HCV evolutionary dynamics during infection. The structural E1 and E2 genes evolve fastest (13) and have the strongest phylogenetic signal (40) as a result of encoding proteins that are recognized by the host immune system. Several studies report that some or most of the pretransplant diversity in the E1E2 region is lost during a viral bottleneck that follows transplantation, possibly reflecting the outgrowth of fitter variants (2, 8, 9, 16, 32, 41, 42). In contrast, some evidence suggests that posttransplant viral dynamics are more complex than a simple population bottleneck; for example, the dominant variant at 7 days posttransplant does not persist in later samples in all cases (2, 42), and the minor variant pretransplant can become dominant (14). An observed bottleneck could be explained by a founder effect of colonization of the new liver, or result from methodological limitations such as consensus sequencing (28) or single-strand conformation polymorphisms (2, 32) and the use of summary statistics (9, 32, 37, 41), which fail to elucidate evolutionary relationships or structure (39, 44); grouping results from multiple patients rendering data interpretation difficult (9, 27, 32, 41, 42); and temporally restric-

Received 27 March 2012 Accepted 18 May 2012
Published ahead of print 23 May 2012
Address correspondence to Rebecca R Gray, rebecca.gray@zoo.ox.ac.uk, or Marco Salemi, salemi@pathology.ufl.edu.

doi:10.1128/JVI.00749-12

TABLE 1 Patient and sample characteristics[a]

Patient code	Age (yr)	Gender	Ethnicity[c]	Biopsy result	Time point (mo.) pre/post-TP date (no. of clones)[b]	Peak RNA level (pre/post-TP) (copies/ml)	Peak ALT lavel (pre/post-TP) (IU/l)
A	51	Male	C	G0S1-2	–32.4 (21), −4.4 (22), 2.3 (24), 12.4 (23)	64,000/>700,000	NA/335
B	57	Male	C	G3S2	–19.0 (23), −0.8 (24), 1.0 (22), 12.2 (22)	374,000/687,000	143/211
C	51	Male	C	G2-3S2	–23.8 (26), −7.7 (24), 1.5 (24), 12.6 (24), 20.1 (24)	47,100/691,000	228/81
D	44	Male	NAm	NA	–18.1 (33), −3.8 (23), 1.6 (19), 12.7 (23), 27.0 (23)	NA/2,510,000	79/40
F	48	Male	AA	G2S0	–24.4 (23), −4.7 (24), 1.5 (24), 14.6 (26), 34.1 (24)	342,000/>700,000	82/139
G	53	Male	C	G2S1	–20.1 (24), 2.3 (23), 6.5 (22)	369,000/1,710,000	146/149
H	54	Male	AA	G2S0	–9.5 (20), −1.1 (19), 3.3 (27), 38.8 (25)	293,000/>5,000,000	76/31
I	48	Male	C	G1S0	–28.1 (23), −2.2 (36), 1.9 (28), 24.9 (26)	426,000/>700,000	85/54
J	53	Male	C	G2S0	–57.1 (25), −10.4 (26), 1.5 (24), 7.6 (25)	584,000/NA	60/36
L	65	Male	C	G2-3S1	–23.4 (24), −1.1 (21), 1.6 (23), 13.2 (24), 23.2 (24)	108,000/4,890,000	65/272

[a] TP, transplantation; NA, not available.
[b] The numbers of clones at each time point are indicated in parentheses.
[c] C, Caucasian; AA, African American; NAm, Native American.

tive sampling schemes that limit investigation of long-term evolutionary trends (8, 9, 16, 37).

We investigated here the long-term evolution of HCV pre- and posttransplant in 10 patients to determine the effect of the transplant on viral dynamics. A longitudinal sampling strategy, spanning at least 1 to 2 years pre- and posttransplant, was combined with high-resolution phylogenetic methods to define the evolutionary relationships in the virus population pre- and posttransplantation. We found that the patterns of reinfection were more complex than previously reported; when appropriate population genetic methods are used, no evidence of a significant viral genetic bottleneck appears following transplantation, and transmitted minor variants can persist for long periods of time. The dynamics of posttransplant reinfection may reflect the poorly understood process of transmission to new hosts with novel immune backgrounds and selective pressures.

MATERIALS AND METHODS

Amplification and sequencing. Serum samples were collected before and after liver transplantation at the University of North Carolina Liver Center under an Institutional Review Board-approved protocol with written informed consent from all participants. Patient characteristics are described in Table 1. Viral RNA was extracted using the QIAamp viral RNA minikit (Qiagen, Valencia, CA). The cDNA synthesis and amplification was performed in a single step, using the Superscript One Step RT-PCR system with Platinum *Taq* DNA polymerase (Life Technologies, Grand Island, NY) according to the manufacturer's conditions with the exception of the primer concentration and thermocycling conditions. For the reaction mix 0.8 μM concentrations of the primers E1_F_3_1_HCV (5′-ATGGCNTGGGAYATGATRATGAA-3′) and E2_R_1_1_HCV (5′-TTCATCCABGTRCARCCRAACCA-3′) were used with 60 ng of RNA template/μl. Also, 0.8 μM concentrations of the primers E1_F_3_1_HCV (5′-ATGGCNTGGGAYATGATRATGAA-3′) and E2_R_2_1_HCV (5′-CCYCGNGTCCARTTGCA-3′) were used for the samples that failed initial amplification. The primer designs and extracted RNA were provided by the Wang laboratory at the University of Florida. The first set spans a 716-bp region, and the second 995-bp region of the *env* region of HCV subtype 1, which includes the hypervariable regions. The thermocycling conditions consisted of 50°C for 30 min and 94°C for 2 min, followed by 35 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1 min, with a final step of 72°C for 10 min. A plasmid containing the full-length sequence of H77C and nontemplate samples were used as controls. The amplicons were run on a 1.5% agarose gel in buffer Tris-acetate-EDTA for confirmation of amplification. Each reaction was performed one time point at a time to prevent cross-contamination. Amplified samples were cloned by using a TOPO TA cloning kit (Life Technologies) according to the manufacturer's conditions. Both directions of the individual clones were sequenced at the University of Florida's Interdisciplinary Center for Biotechnology Research facility using the TOP10-specific M13 forward primer and the M13 reverse primer. All sequences were assembled using CodonCode software (CodonCode Corp., Dedham, MA). The analyzed sequences spanned from nucleotides 1316 to 1984 relative to the reference H77 genome (accession number NC_004102).

Nucleotide diversity. Serum HCV sequences were obtained for three to five time points surrounding surgery, ranging from 1 to 12 months and 2 years pretransplant, and within 4 months, 6 to 14 months, and >20 months posttransplant. For each patient and time point, an average of 24 cloned sequences was obtained (range, 19 to 33 sequences). The sequences were separately aligned manually using BioEdit for each patient. The nucleotide diversity at each time point (mean pairwise genetic distances) was calculated separately using a maximum-likelihood correction and gamma-distributed rate variation among sites. Standard errors were calculated using 200 bootstrapped replicates. These analyses were performed in MEGA v.4 (46).

Bayesian rate and coalescent growth models. Bayesian phylogenetic analyses were performed using the HKY+G model of nucleotide substitution, with independent estimates at the codon 1+2 and 3 positions, implemented in the program BEAST v.1.6 (4). Convergence was assessed by visual examination of traces and ensuring all effective sampling size (ESS) values of >200 in the program Tracer v.1.5. Initially, a constant size coalescent prior with the strict molecular clock was assumed. This was then compared to a model that used a relaxed molecular clock that allows uncorrelated rate variation among lineages. Both of these models assumed constant population diversity through time.

The constant diversity assumption was relaxed and a "two-epoch" model was used (7), with the date of the transplant used as a strong prior for the transition time between epochs. All parameters were held constant across the two epochs except for the population size parameters. Two different epoch models were tested: one in which the population diversity was constant in each of the epochs and one in which the population diversity increased exponentially.

Finally, the Bayesian skyline plot model was used; this model allows the population diversity to change throughout time (5, 35). The marginal likelihoods of competing models were compared using Bayes factors. Strong evidence in favor of the alternative model is provided when 2(lnL model 1 − lnL model 2) is >10 (45).

Estimation of new lineages. In order to estimate the number of lineages that emerged posttransplant, we counted the number of changes between two phylogenetic states (pre- and posttransplant). Sequences

were coded as one of the two states, and internal nodes states were reconstructed using parsimony with an irreversible matrix, such that only pre- to posttransplant state changes were permitted (as implemented in MacClade). For each patient, the mean and 95% confidence intervals (95% CI) of the number of emergent lineages were calculated from the distribution of changes across the posterior distribution of trees (minus a 10% burn-in).

Maximum-likelihood trees. Maximum-likelihood phylogenies were inferred in PhyML (15) using a general time reversible (GTR) model of nucleotide evolution and gamma-distributed rate variation among sites, with no molecular clock enforced. Each patient was analyzed separately. with all time points included. A total of 200 bootstrap replicates were performed to assess the statistical support for the topologies. Trees were visualized and annotated in FigTree v.1.3 and rooted according to the "best root" option in Pathogen v.1.2 (http://tree.bio.ed.ac.uk/software/pathogen).

Selection analysis. Alignments from each of the 10 patients were tested using the single-likelihood ancestor counting (SLAC) and fixed-effects likelihood (FEL) methods to detect site-specific selection via an online server (http://www.datamonkey.org [33]). These methods were chosen for their applicability to large (>50 taxa) data sets and short sequences. The best-fitting model of nucleotide substitution was used. An alpha value of 0.05 was used to assess significance.

Nucleotide sequence accession numbers. Nucleotide sequences for the HCV strains have been deposited in GenBank under accession numbers JQ063473 to JQ064506.

RESULTS

Clinical cohort. Ten patients included in the study were male and between the ages of 44 and 57 (Table 1). In all but one case (patient B) the patients were alive at the time of the study. Serum samples were obtained for three to five time points surrounding surgery, ranging from 1 to 12 months and >12 months pretransplant and <4 months, 6 to 14 months, and >20 months posttransplant. A liver biopsy performed after the transplant indicated variable progression to fibrosis with all patients showing mild to moderate progression ($G \leq 3$).

Nucleotide diversity. First, to evaluate the expectation of a sharp decrease in diversity posttransplant under the bottleneck model, sample diversities were calculated as the mean pairwise genetic distance among sampled sequences for each patient/time point without taking into account phylogenetic relationships among sequences (Fig. 1). Sample diversities across patients were variable, ranging from <0.01 substitutions/site (patient G) to > 0.07 substitutions/site (patient A). No consistent trend in decreasing genetic diversity during the initial 4 months posttransplant was apparent. In patient F, diversity significantly increased in the first sample posttransplant relative to the diversity present 5 months pretransplant. No significant change in sample diversity at the first time point posttransplant was found in six patients, while a significant decrease was detected in three patients (A, D, and I).

Over longer periods of time (6 to 15 months posttransplant), viral diversity showed no trend across patients. A statistically significant increase in diversity appeared in patients A and D relative to the first sample posttransplant, while a statistically significant decrease was observed for patients J, B, and F. In the six patients with a sample > 20 months posttransplant (D, I, C, F, H, and L), two (D and H) exhibited a statistically significant increase in genetic diversity relative to the previous samples. In patient F, sample diversity 34 months posttransplant was similar to diversity in the sample 1 month posttransplant, even though a transient reduction in diversity appeared in the sample taken 14 months posttransplant. In two patients (C and L) a significant decrease in diversity was observed by 20 and 23 months posttransplant, respectively, compared to the previous time point.

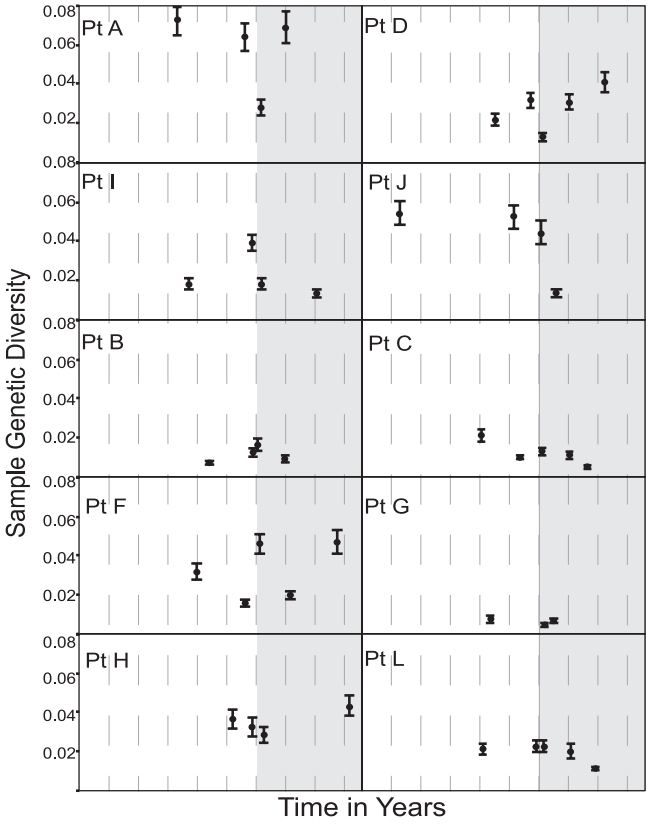

FIG 1 Sample genetic diversity in serum. Plots are shown for patients A to L. Sample genetic viral diversity was assessed for all samples in 10 HCV patients using maximum-likelihood corrected pairwise nucleotide distances for all sequences at a given time point. For each patient, the *y* axis shows estimated genetic diversity. The *x* axis represents time going forward. Each of the light gray vertical bars = 1 year. The gray shaded area represents the time after the transplant. Mean and standard error are denoted with a circle and bars.

Overall, evidence for a bottleneck model using the common approach of mean pairwise diversity was identified in only three of the 10 subjects (A, D, and I), while in seven subjects stable or increased posttransplant diversity supported an alternative explanation.

Coalescent growth models. Several evolutionary models that incorporate phylogenetic relationships among sequences were considered to quantify changes in the diversity of the whole viral population over time. First, a strict molecular clock model (all lineages evolve at the same rate) was compared against a more general relaxed clock model (which allows lineages to evolve at different rates). By comparing the marginal likelihoods of the two models using Bayes factors (BFs), the relaxed clock model was strongly favored over the simpler strict clock for all 10 patients (BF > 50), providing a rational for using the relaxed clock model in all subsequent analyses.

Next, population genetic diversity through time for each patient was estimated using a Bayesian skyline plot (BSP) model, which includes estimates of statistical uncertainty (5). Under a bottleneck scenario, a sharp reduction in the relative population

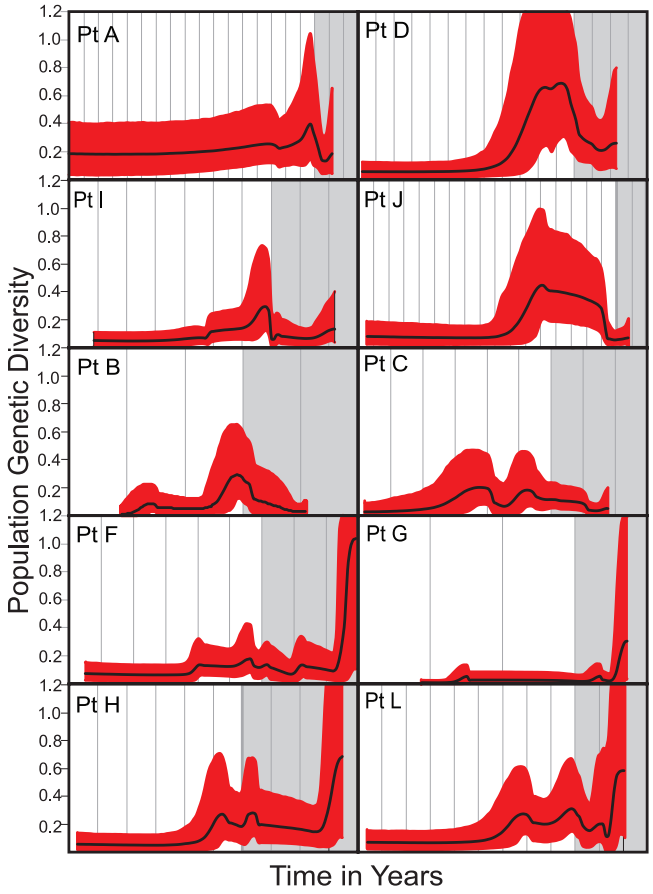

FIG 2 Population genetic diversity. Plots are shown for patients A to L. For each patient, the *y* axis shows estimated genetic diversity. The *x* axis represents time going forward. Each of the vertical bars = 1 year. The gray shaded area represents the time after the transplant. The black curve gives the mean estimate of the BSP, and the red area indicates the 95% high posterior density intervals.

genetic diversity around the time of transplantation would be expected if a limited number of founder viruses successfully establish infection in the new liver, whether resulting from stochastic or deterministic forces. BSPs showed no significant reduction in population genetic diversity at the time of the transplant across patients (Fig. 2). Most patients showed constant population diversity throughout infection, suggesting that the transplant did little to disrupt the overall diversity of the population. Although five patients (A, D, I, J, and B) showed a slight decline in HCV population diversity around the time of the transplant, the decrease was well within the wide confidence intervals and thus not significant. We then compared the fit of the general BSP model to a much simpler model of constant population diversity; only for patient B was the BSP strongly favored (BF $>$ 10) over the simpler model. Interestingly, patient B was not one of the patients in the previous analysis identified as displaying a putative bottleneck. Thus, in 9/10 cases a model of constant population diversity cannot be rejected.

We next tested two-epoch models. First, viral populations were allowed to have one level of genetic diversity before transplant and another level afterwards (white and gray areas, respectively, in Fig. 2). Alternatively, the posttransplant diversity was allowed to increase exponentially. Neither epoch model was favored over the simplest model of constant diversity through time (BF $<$ 10). The results are consistent with those from the BSPs and overall suggest that the viral population diversity remained unchanged after the transplant.

Phylogenetic analysis. To understand the evolutionary history of the posttransplant virus, maximum clade credibility genealogies assuming a relaxed molecular clock and constant population diversity were inferred. The expectation under a bottleneck scenario was that only one lineage, representing the founder virus that survived the bottleneck and initiated infection in the new liver, would emerge after transplantation and that all subsequent posttransplant viruses would share a common ancestor with this founder strain.

The bottleneck signature was observed for patient G, in which the vast majority of sequences sampled within 4 months posttransplant clustered together (Fig. 3). In contrast, multiple origins of the posttransplant viral population were observed for 9 of 10 patients. Unexpectedly, viruses sampled within 4 months posttransplant often shared a common ancestor with viruses from 2 or more years pretransplant, rather than the time point immediately pretransplant. For example, in patients B and F, a major lineage containing sequences from 1 to 2 months posttransplant, respectively, shared an ancestor with viruses sampled at 19 and 24 months posttransplant, creating a distinct clade that includes no sequences sampled 1 and 5 months posttransplant (bottom clade in both patients in Fig. 3).

In several cases (e.g., patients J, B, and G), sequences sampled within 4 months posttransplant gave rise to the majority of lineages present in the sample 6 to 15 months posttransplant. Unexpectedly, in other cases some lineages sampled 6 to 15 months posttransplant did not share a origin with the 0- to 4-month posttransplant sample but rather with an earlier sample (e.g., patients A, D, C, and L). In some of these cases, the origin of the 6- to 15-month posttransplant sample was quite close to the root (e.g., patients A and D), indicating that the lineage was in existence—but not detected—for a period of years.

Of the six patients with a sample $>$20 months posttransplant (D, I, C, F, H, and L), four (D, I, F, and H) showed multiple lineages containing sequences from this last time point. Interestingly, again, some of these clades did not share an ancestor with the immediately preceding time point, but rather one further back in time. In patient D, most of the sequences from 27 months posttransplant shared an ancestor with the pretransplant viruses pretransplant, in a clade that contained no sequences from 1 or 12 months posttransplant. In patient H, the major clades containing the 38 month posttransplant sequences did not contain any sequences from 3 months posttransplant but rather exclusively pretransplant sequences. The other two patients (C and L) had a single clade containing sequences sampled $>$20 months posttransplant. We used a phylogenetic method to estimate the number of viral lineages that survive the transplantation process (Table 2). The mean number of surviving lineages summed over all trees in the posterior distribution (the same as used to generate the MCC tree, above) ranged from 1 (patient G) to 15 (patients A and D). In only one instance did the distribution of the estimated number of re-emerging lineages include 1 (patient G), providing strong evidence for multiple posttransplant founder lineages. Maximum-likelihood trees showed a similar pattern (data not shown).

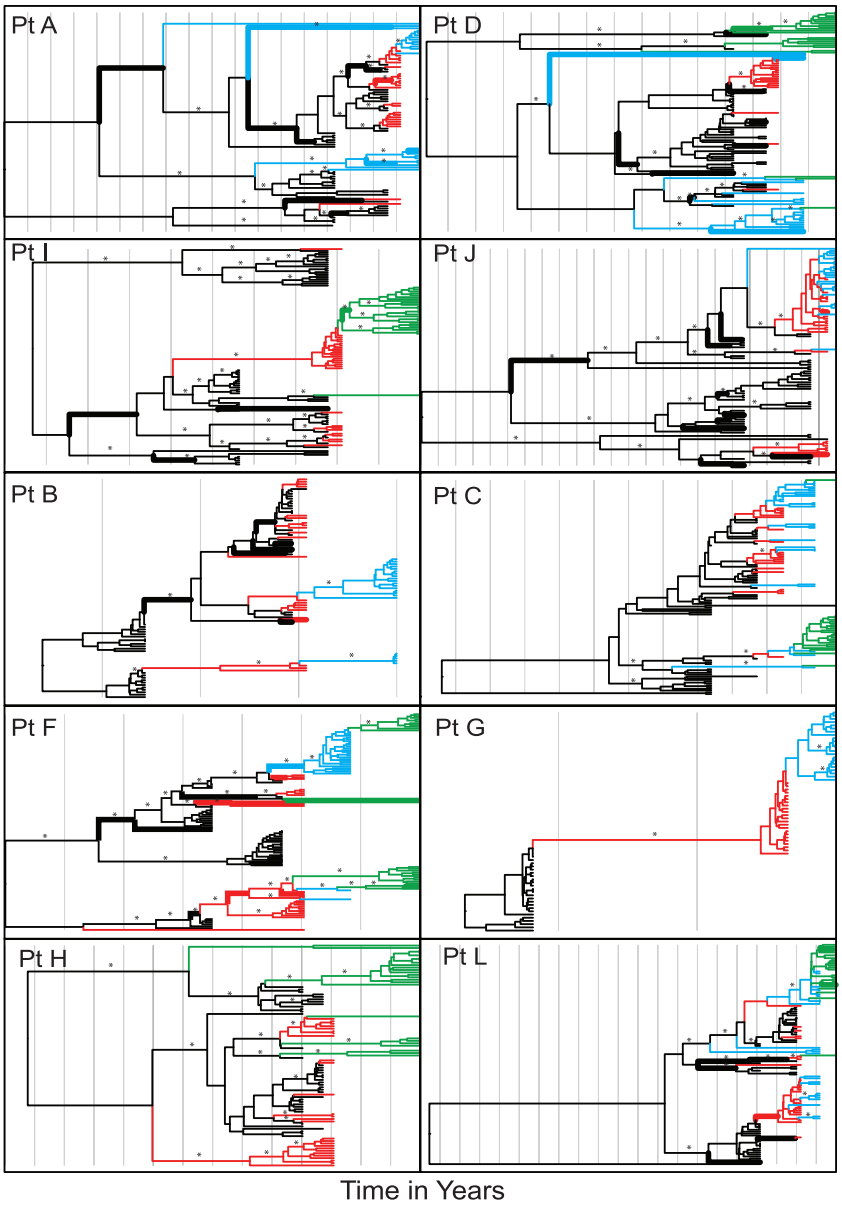

FIG 3 Bayesian phylogenies. Trees are shown for patients A to L where branches are scaled in time with vertical lines indicating years (drawn relative to the last sample). Terminal branches are colored according to the time of sampling: black = pretransplant, red = 1 to 3 months posttransplant, blue = 3 to 20 months posttransplant, green = >20 months posttransplant. Internal branches are colored according to the maximum-parsimony reconstruction of the time of sampling with a time-forward enforcement. Asterisks indicate posterior support > 0.9, and thick branches indicate a lineage containing a positively selected site.

TABLE 2 Number of distinct posttransplant lineages

Patient	Mean no. of transmitted lineages (95% CI)[a]
A	14.85 (11–21)
B	11.55 (7–17)
C	11.17 (6–17)
D	14.72 (11–23)
F	5.85 (5–8)
G	1 (1–1)
H	10.21 (6–15)
J	8.65 (6–12)
K	6.28 (5–7)
L	9.73 (8–13)

[a] 95% CI, 95% confidence interval.

Overall, these results demonstrate that a wide section of pretransplant diversity is maintained through the posttransplant infection. In some cases, the dominant major variant at a later posttransplant time point appears to have evolved from a lineage present at a level below detection (~5%) for many months or even years. Only patient G showed the expected pattern of a bottleneck scenario in the phylogenetic tree; however, note that this patient is unique in only having three time points represented. These results contrast with the apparent signal of a bottleneck for patients A, D, and I in the sample diversity analysis, which demonstrates the importance in incorporating phylogenetic relationships when testing such hypotheses (39).

Selection analysis. For each patient, we used two methods—

TABLE 3 Site-specific selection[a]

Patient	Site selection					
	SLAC			FEL		
	Site	*dN/dS*	*P*	Site	*dN/dS*	*P*
A	330	7.35	0.04	330	3.90	0.05
	405	9.92	0.01	**399**	4.79	0.03
	498	9.01	0.02	**404**	3.43	0.03
				405	6.07	0.00
				498	5.43	0.01
				538	2.87	0.05
B	**397**	16.06	0.02	**397**	35.06	0.01
				477	14.55	0.02
C	None			None		
D	*533*	6.31	0.02	*532*	6.06	0.02
F	**405**	9.69	0.04	334	7.16	0.01
				400	4.51	0.02
				405	5.05	0.01
G	None			None		
H	None			None		
I	None			445	7.81	0.04
J	445	4.99	0.04	**386**	2.32	0.05
	446	8.25	0.01	**395**	3.92	0.04
				405	2.45	0.03
				445	5.78	0.01
				446	8.40	0.03
L	None			**384**	3.31	0.03
				394	3.42	0.05
				401	2.23	0.03

[a] Numbering is indicated relative to the H77 reference genome. Boldfacing indicates a position within HVRI (amino acids 382 to 410); italics indicate a position in a defined epitope (amino acids 425 to 443 and 529 to 535).

SLAC and FEL—for detecting site-specific positive selection (Table 3). In 7 of 10 patients, at least one site was detected as significantly under positive selection ($P < 0.05$). In general, FEL identified additional sites over SLAC, which is consistent with the conservative nature of the SLAC test (20). In one case (patient D), FEL failed to detect a site that was identified by SLAC. In total, 22 different sites across the seven patients were detected as being under positive selection. The majority (12/22) were in the hypervariable region (HVRI) region of E2 (amino acids 384 to 410 relative to H77 reference genome), which contains epitopes for neutralizing antibodies and may also act as an epitope for helper (43) and cytotoxic (48) T cells. The two sites identified for patient B were localized in epitopes for monoclonal antibodies that contain contact residues for CD81 cellular receptor binding (18, 19).

Changes at these sites were then mapped onto the specific lineages on the MCC trees (Fig. 3). In general, positive selection was present throughout the history of the infection, indicating ongoing adaptive evolution. In several cases (patients D, I, F, and L), these sites were found on posttransplant lineages that led to one of the majority variants.

DISCUSSION

Our study revealed that multiple HCV lineages are transmitted at the time of liver transplant without a major decrease in viral genetic diversity. Although only some of the pretransplant lineages are identified within the first 4 months posttransplant, lineages are undoubtedly present because their ancestors are sampled at later time points. The data clearly argue against a bottleneck scenario in most transplant recipients and suggest that little restriction of diversity occurs in the new liver in which multiple lineages set up a new infection.

It could be suggested that a restriction in the viral population is unlikely to be observed in the 4 months posttransplant, since insufficient time has elapsed and patients are typically on immunosuppressive drugs during this period that may reduce the selective pressure on the virus. However, we show the absence of a transplantation bottleneck for as long as 1 to 2 years posttransplant. Although analysis of sample genetic diversity was consistent with a bottleneck in 3/10 patients (A, D, and I), subsequent population genetic analyses plainly demonstrated by multiple parameters that the genetic bottleneck signature failed to apply to the virus populations in these patients and that estimates based on sample diversities have underestimated the actual diversity of the viral population.

Furthermore, multiple distinct viral lineages sampled a year or more after transplant share a ancestor with viruses sampled well before (2 years) transplant, rather than with viruses sampled within 4 months posttransplant, as would be expected for a continuously replicating virus. Our sequencing strategy allowed for detection of viral variants at ~5% level (average of 24 clones/time point). Thus, it is possible—and indeed very likely—that variants belonging to these unobserved lineages were present in the serum at levels below our threshold of detection. However, a fundamental tenet arising from molecular evolutionary theory is that, in the absence of a active mechanism of maintenance, segregating variants in a population will be lost, either due to fixation (via random genetic drift or positive selection) or due to elimination (via genetic drift or negative selection). The mean survival time of very rare segregating variants (whether advantageous, neutral, or deleterious) is particularly short as a result of their high probability of stochastic extinction (see, for example, reference 38). Thus, the persistence of rare variants over long periods of time requires an active evolutionary force, such as frequency-dependent selection (whereby the fitness of a variant declines as its frequency increases) or spatially conditioned selective pressure that results in local adaptations (and thus population structure) (31). Our present study cannot distinguish between these two possibilities. In the latter, the population structure could arise from spatial segregation in the liver or from virus replication in additional anatomic compartments (39). For example, hepatic lymph nodes (37) and peripheral blood mononuclear cells (25) have been suggested possible extra-hepatic sources of infection for the new liver. Other possible reservoirs include macrophages (23, 24, 36), the central nervous system (10, 22), and/or B cells (6). However, previous models have suggested that only a small fraction of posttransplant viruses originate from these sources (34).

Primer/PCR bias is unlikely to be a factor in the detection of the major variants, as a bias would differentially affect the pool of variants such that one particular type would be preferentially amplified. In the present data set, however, entirely different variants were amplified at various time points. PCR misincorporation errors and/or recombination would not account for the deep lineage structure observed. Interestingly, an ongoing independent analysis of the clinical samples studied here using pyrosequencing of the NS3 region with a limit of detection of ~1% suggests a similar pattern in which clades contain sequences from well before and after transplant, but none from the time points surrounding the

operation (G. Wang, unpublished data). This observation will be investigated further in future work.

A complex suite of selective pressure is likely to operate on the virus in this cohort, including the potentially strong selective pressure of a new liver expressing different HLA alleles, as well as a potential reduction in host immune pressure as a result of immunosuppressive therapy. Positively selected sites were detected in 7 of 10 patients, mainly in the HVRI region. These sites are most likely to be under diversifying selection as methods using the *dN/dS* test within a population are the most sensitive to this type of selection, and this region is known to be the target of T and B cell immune responses. Lineages containing sites under positive selection were distributed through time and in several cases led to one of the dominant variants posttransplant, indicating a potential selective advantage of those variants.

The results of the present study differ significantly from those reported previously. Our study included a longer temporal sampling strategy and used more extensive analysis methods. Although mean pairwise genetic distance measures of *sample* diversity are one method for assessing bottlenecks, the statistic is readily confounded when the sampled population is structured into subpopulations (49) because pairwise distance ignores the phylogenetic relationships among samples. In contrast, measures of population diversity explicitly take in account the phylogenetic relationships among samples and can infer the presence of unsampled lineages at earlier time points from the observation of their descendants at later times. In addition, a long-term sampling strategy that included viruses sampled from several years before and after the transplant revealed the nature and complexity of the viral population that reinfects the new liver, which would have been obscured by a simple cross-sectional analysis immediately after the transplant.

All virus populations in the present study were obtained from serum. Although such viruses are often assumed to represent the viral population in the liver, serum viruses may also contain variants from nonhepatic sites. Future studies should include both liver biopsies and long-term temporal sampling to understand the dynamics of transplant reinfection.

ACKNOWLEDGMENTS

We thank Gary Wang for helpful discussions.

R.R.G. was funded by an NIH Training Grant in Cell Biology and Immunobiology of Cancer (T32 CA912629), a University of Florida Pathology EPIG award, and currently by the UK Medical Research Council. M.W.F. is funded, in part, by K24 DK066144. O.G.P. is supported by the Royal Society. M.M.G. is funded, in part, by the Stephany W. Holloway University Chair for AIDS Research. D.R.N. is funded, in part, by K24 CA139570. This study was supported in part by the NIH/NCRR Clinical and Translational Science Award to the University of Florida UL1 RR029890.

REFERENCES

1. **Allain JP, Dong Y, Vandamme AM, Moulton V, Salemi M.** 2000. Evolutionary rate and genetic drift of hepatitis C virus are not correlated with the host immune response: studies of infected donor-recipient clusters. J. Virol. **74**:2541–2549.
2. **Arenas JI, et al.** 2004. Hepatitis C virus quasi-species dynamics predict progression of fibrosis after liver transplantation. J. Infect. Dis. **189**:2037–2046.
3. **Centers for Disease Control and Prevention.** 2010. Hepatitis C FAQs for the public. Centers for Disease Control and Prevention, Atlanta, GA. http://www.cdc.gov/hepatitis/C/cFAQ.htm.
4. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7**:214. doi:10.1186/1471-2148-7-214.
5. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. **22**:1185–1192.
6. **Ducoulombier D, et al.** 2004. Frequent compartmentalization of hepatitis C virus variants in circulating B cells and monocytes. Hepatology **39**:817–825.
7. **Edwards CT, et al.** 2006. Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. BMC Evol. Biol. **6**:28. doi:10.1186/1471-2148-6-28.
8. **Fafi-Kremer S, et al.** 2010. Viral entry and escape from antibody-mediated neutralization influence hepatitis C virus reinfection in liver transplantation. J. Exp. Med. **207**:2019–2031.
9. **Feliu A, Gay E, Garcia-Retortillo M, Saiz JC, Forns X.** 2004. Evolution of hepatitis C virus quasispecies immediately following liver transplantation. Liver Transpl. **10**:1131–1139.
10. **Fletcher NF, et al.** 2012. Hepatitis C virus infects the endothelial cells of the blood-brain barrier. Gastroenterology **142**:634–643.e636.
11. **Forman LM, Lewis JD, Berlin JA, Feldman HI, Lucey MR.** 2002. The association between hepatitis C infection and survival after orthotopic liver transplantation. Gastroenterology **122**:889–896.
12. **Gane EJ, et al.** 1996. A longitudinal analysis of hepatitis C virus replication following liver transplantation. Gastroenterology **110**:167–177.
13. **Gray RR, et al.** 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evol. Biol. **11**:131. doi:10.1186/1471-2148-11-131.
14. **Gretch DR, et al.** 1996. Tracking hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. J. Virol. **70**:7622–7631.
15. **Guindon S, Delsuc F, Dufayard J, Gascuel O.** 2009. Estimating maximum-likelihood phylogenies with PhyML. Methods Mol. Biol. **537**:113–137.
16. **Hughes MG, Jr, et al.** 2004. E2 quasispecies specificity of hepatitis C virus association with allografts immediately after liver transplantation. Liver Transpl. **10**:208–216.
17. **Ina Y, Mizokami M, Ohba K, Gojobori T.** 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. J. Mol. Evol. **38**:50–56.
18. **Keck ZY, et al.** 2008. Definition of a conserved immunodominant domain on hepatitis C virus E2 glycoprotein by neutralizing human monoclonal antibodies. J. Virol. **82**:6061–6066.
19. **Keck ZY, et al.** 2011. Mapping a region of hepatitis C virus E2 that is responsible for escape from neutralizing antibodies and a core CD81-binding region that does not tolerate neutralization escape mutations. J. Virol. **85**:10451–10463.
20. **Kosakovsky-Pond SL, Frost SD.** 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. **22**:1208–1222.
21. **Kuntzen T, et al.** 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Hepatology **48**:1769–1778.
22. **Laskus T, et al.** 2002. Detection and analysis of hepatitis C virus sequences in cerebrospinal fluid. J. Virol. **76**:10064–10068.
23. **Laskus T, et al.** 2004. Human immunodeficiency virus facilitates infection/replication of hepatitis C virus in native human macrophages. Blood **103**:3854–3859.
24. **Laskus T, et al.** 2000. Hepatitis C virus in lymphoid cells of patients coinfected with human immunodeficiency virus type 1: evidence of active replication in monocytes/macrophages and lymphocytes. J. Infect. Dis. **181**:442–448.
25. **Laskus T, Radkowski M, Wilkinson J, Vargas H, Rakela J.** 2002. The origin of hepatitis C virus reinfecting transplanted livers: serum-derived versus peripheral blood mononuclear cell-derived virus. J. Infect. Dis. **185**:417–421.
26. **Levrero M.** 2006. Viral hepatitis and liver cancer: the case of hepatitis C. Oncogene **25**:3834–3847.
27. **Li H, et al.** 2010. Genetic diversity of hepatitis C virus predicts recurrent disease after liver transplantation. Virology **402**:248–255.
28. **Massaguer A, et al.** 2007. Evolution of the NS3 and NS5B regions of the hepatitis C virus during disease recurrence after liver transplantation. Am. J. Transplant. **7**:2172–2179.

29. **McHutchison JG, et al.** 2009. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N. Engl. J. Med. **360**:1827–1838.
30. **Meuleman P, Leroux-Roels G.** 2009. HCV animal models: a journey of more than 30 years. Viruses **1**:222–240.
31. **Mitchell-Olds T, Willis JH, Goldstein DB.** 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. **8**:845–856.
32. **Moreno-Garcia J, et al.** 2003. Analysis of hepatitis C viral quasispecies in liver transplantation. Transplant Proc. **35**:1838–1840.
33. **Pond S, Frost S.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21**:2531–2533.
34. **Powers KA, et al.** 2006. Kinetics of hepatitis C virus reinfection after liver transplantation. Liver Transpl. **12**:207–216.
35. **Pybus OG, Rambaut A, Harvey PH.** 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics **155**:1429–1437.
36. **Radkowski M, et al.** 2004. Infection of primary human macrophages with hepatitis C virus in vitro: induction of tumour necrosis factor-alpha and interleukin 8. J. Gen. Virol. **85**:47–59.
37. **Ramirez S, et al.** 2009. Hepatitis C virus compartmentalization and infection recurrence after liver transplantation. Am. J. Transplant. **9**:1591–1601.
38. **Rouzine IM, Coffin JM.** 1999. Linkage disequilibrium test implies a large effective population number for HIV in vivo. Proc. Natl. Acad. Sci. U. S. A. **96**:10758–10763.
39. **Gray RR, Salemi M, Klenerman P, Pybus OG.** 2012. A new evolutionary model for hepatitis C virus chronic infection. PLoS Pathog. **8**:e1002656. doi:10.1371/journal.ppat.1002656.
40. **Salemi M, Vandamme A.** 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. J. Mol. Evol. **54**:62–70.
41. **Schramm F, et al.** 2008. Frequent compartmentalization of hepatitis C virus with leukocyte-related amino acids in the setting of liver transplantation. J. Infect. Dis. **198**:1656–1666.
42. **Schvoerer E, et al.** 2007. Early evolution of hepatitis C virus (HCV) quasispecies after liver transplant for HCV-related disease. J. Infect. Dis. **196**:528–536.
43. **Shirai M, et al.** 1999. T cell recognition of hypervariable region-1 from hepatitis C virus envelope protein with multiple class II MHC molecules in mice and humans: preferential help for induction of antibodies to the hypervariable region. J. Immunol. **162**:568–576.
44. **Steel MA, Hendy MD, Penny D.** 1988. Loss of information in genetic distances. Nature **336**:118.
45. **Suchard MA, Weiss RE, Sincheimer JS.** 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. **18**: 1001–1013.
46. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.
47. **Thomas E, Fried MW.** 2008. Hepatitis C: current options for nonresponders to peginterferon and ribavirin. Curr. Gastroenterol. Rep. **10**: 53–59.
48. **Urbani S, et al.** 2001. Identification of immunodominant hepatitis C virus (HCV)-specific cytotoxic T-cell epitopes by stimulation with endogenously synthesized HCV antigens. Hepatology **33**:1533–1543.
49. **Wang X, et al.** 2007. Progression of fibrosis during chronic hepatitis C is associated with rapid virus evolution. J. Virol. **81**:6513–6522.
50. **World Health Organization.** 2010. HCV: surveillance and control. World Health Organization, Geneva, Switzerland. http://www.who.int/csr/disease/hepatitis/whocdscsrlyo2003/en/index4.html.